

CAI, Congreso Argentino de AgroInformática

Mapecto de cultivos agrícolas en Córdoba: una comparación del desempeño de tres algoritmos de clasificación

Miguel Nolasco¹, Mónica Bocco¹ y Carlos Scavuzzo²¹ Facultad de Ciencias Agropecuarias, Universidad Nacional de Córdoba. Argentina.² Comisión Nacional de Actividades Espaciales, Argentina.mnolasco@agro.unc.edu.ar

Resumen. El uso de sensores remotos permite el mapeo y monitoreo de cultivos en grandes superficies. La capacidad de identificación por especie y en tiempo casi real es deseable. En el presente trabajo se describe el uso práctico de la teledetección en el mapeo de cultivos. El objetivo del estudio fue comparar el desempeño de los algoritmos de clasificación Máxima Verosimilitud, Support Vector Machine, Random Forest, y analizar el efecto de la naturaleza y número de clases a entrar en los modelos. Los resultados manifiestan la posibilidad de mapear la diversidad de coberturas agrícolas presentes, con anterioridad a la finalización de sus ciclos. Las mejores índices de precisión se alcanzaron al utilizar grupos de entrenamiento-validación, cuyas clases representaban los estadios fenológicos de cada cultivo. Se respalda la conveniencia del algoritmo de Máxima Verosimilitud, el cual alcanzo excelentes resultados, fue de ejecución simple y de bajo costo computacional.

Palabras clave: Máxima verosimilitud, Support Vector Machine, Random Forest.

1 Introducción

El mapeo de cultivos por medio de técnicas de teledetección posee ventajas de alta eficiencia y bajo costo, comparado con métodos basados en encuestas, lo cual hace factible la generación de cartografía en la mayoría de los casos [1]. Los métodos basados en encuestas a terreno pueden generar mapas más precisos en comparación a los basados en teledetección. Sin embargo si la tasa de cambio de uso del suelo es considerable, el enfoque basado en teledetección con capacidad de mapeo más frecuente, es necesario [2]. Para la planificación de políticas agropecuarias a corto plazo, la capacidad de identificación del tipo de cultivo en tiempo real o casi en tiempo real es deseable, ya que, por ejemplo permitiría proporcionar una estimación actual de la producción y decidir sobre consecuentes subvenciones [3]. Por lo tanto se requieren algoritmos y procedimientos metodológicos que nos permitan generar mapas por tipo de cultivo en momentos tan cercanos al comienzo de la temporada de crecimiento como sea posible.

A fin de realizar un mapeo de las superficies dedicadas a la agricultura extensiva, a partir de recursos de la teledetección, diversos autores evaluaron distintos algoritmos

de clasificación experimentados. El algoritmo de la clasificación de Máxima Verosimilitud (MLC), es un enfoque de clasificación supervisada ampliamente utilizado, con ventajas tales como: i) estar disponible en la mayoría de entornos de análisis de datos, ii) ser un método relativamente robusto poco probable de producir resultados anormales o caer en el sobre ajuste, y iii) generar salidas predecibles y comprensibles al ser un clasificador supervisado paramétrico[2]. También se han explorado las potencialidades de los algoritmos pertenecientes al campo del aprendizaje automático: Support Vector Machine [4] y Random Forest [5].

En el presente trabajo los objetivos fueron: i) determinar la conveniencia de datos espectrales de fin de temporada para decidir el tipo de cobertura de suelo agrícola en la región central de Córdoba; ii) evaluar y comparar el desempeño de los algoritmos de clasificación de Máxima Verosimilitud, SVM y RF para su detección y iii) analizar el efecto de la naturaleza y número de clases en los distintos modelos.

2 Materiales y métodos

2.1 Área de estudio

El presente trabajo fue realizado en la zona central de la provincia de Córdoba, departamentos Río Segundo y Tercero Arriba ubicación aproximada 31,8°S 63,8°W (Figura 1). Alrededor de este punto central se realizó un recorte de forma rectangular sobre las imágenes a procesar, de aproximadamente 2000 kilómetros cuadrados de área. La precipitación anual promedio en esta región es de 800 mm, concentrándose en los meses de verano. La temperatura media anual es de 16°C, con una amplitud térmica de 14°C [6]. La vegetación natural dominante es el bosque xerófilo, alternando con estepas de gramíneas duras, en el cual se observa un acentuado proceso de desmonte [7]. La fisiografía del área responde a la Pampa loésica alta, con una suave pendiente regional hacia el este. Los suelos son clasificados como Haplustoles énticos y típicos, bien drenados de textura franco limosa [8]. El método de labranza predominante es el de siembra directa, con los cultivos de soja (*Glycine max* (L.) Merrill), maíz (*Zea mays* L.), sorgo (*Sorghum bicolor* (L.) Moench) y maní (*Arachis hypogaea* L.), como predominantes en la época estival [9].

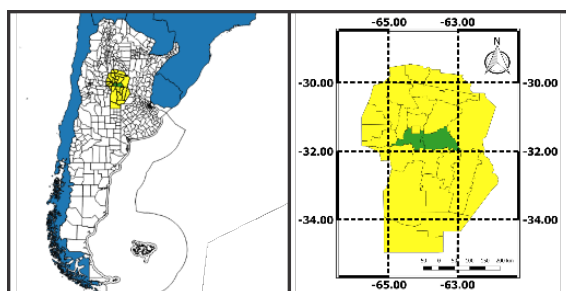


Figura 1. Localización del área de estudio, departamentos Río Segundo y Tercero Arriba resaltados en color verde.

2.2 Datos de campo

Los datos de referencia sobre tipo de cobertura agrícola fueron relevados a campo durante la campaña agrícola 2015-2016. En total se inspeccionaron 70 lotes de uso agrícola. Las visitas a campo se realizaron en los meses de enero y marzo de 2016, registrándose cultivo presente, cultivo antecesor (mediante análisis del rastrojo remanente), estado sanitario y etapa fenológica del cultivo. Sobre esta última característica es importante realizar un análisis exhaustivo, ya que son datos claves que permiten diferenciar la estrategia respecto a la fecha de siembra elegida por el productor. Poder establecer tipos de cobertura según la ubicación temporal del ciclo del cultivo, permite diseñar una correcta metodología de clasificación de áreas dedicadas a la agricultura. De la totalidad de registros, aproximadamente la mitad se utilizó en el entrenamiento de los algoritmos y la otra mitad en la etapa de validación.

En la visita a campo (marzo 2016) se registraron en los distintos cultivos los estadios fenológicos que permitieron establecer las siguientes clases de coberturas:

- C1: Soja temprana (VE a R3).
- C2: Soja intermedia (R4 a R6).
- C3: Soja tardía (R7 a R8).
- C4: Maíz extra temprano (VE a VT).
- C5: Maíz temprano (R1 y R2).
- C6: Maíz intermedio (R3).
- C7: Maíz tardío (R4 y R5).
- C8: Maíz extra tardío (R6).
- C9: Sorgo
- C10: Maní

A partir de las mismas se constituyeron tres grupos, que abarcaban las clases de cobertura con diferentes criterios.

El primer grupo, G1, incluyó las clases correspondientes a los cuatro cultivos en sus diversos estadios fenológicos de manera independiente. Es decir se ingresaron al modelo las diez clases de cobertura.

El segundo grupo, G2, solo considero las clases pertenecientes a los cultivos de soja y maíz, en sus respectivos estadios de ciclo de cultivo. Es decir se ingresaron las clases: C1, C2, C3, C4, C5, C6, C7, y C8.

El tercer grupo, G3, abarco únicamente dos clases: soja y maíz. Para ello se unieron respectivamente las clases que representaban los distintos estadios fenológicos de un mismo cultivo.

2.3 Datos satelitales y procesamiento

Una imagen Landsat 8 del área de estudio (path/row: 229/82) cuya fecha de adquisición se corresponde con el momento en el cual los cultivos se encontraban en etapas

reproductivas de su ciclo (marzo 2016), fue descargada del sitio web del Servicio Geológico de los Estados Unidos (<http://earthexplorer.usgs.gov/>). El tiempo era bueno para el periodo de adquisición de los datos Landsat, las nubes estaban casi ausentes y no se presentaba smog en la atmósfera.

El procesamiento de los datos del sensor OLI incluyó el reescalado a valores de reflectancia en el techo de la atmósfera TOA, usando para ello los coeficientes radiométricos provistos en el archivo de metadato MTL.txt. La siguiente ecuación fue usada para convertir los niveles digitales (ND) obtenidos por el sensor OLI a valores de reflectancia [10]:

$$P\lambda' = M_p Q_{cal} + A_p \quad (1)$$

Donde:

$P\lambda'$: Es el valor de reflectancia planetaria, sin corrección por ángulo solar.

M_p : Es el factor multiplicativo de escalado específico por banda, obtenido del metadato (REFLECTANCE_MULT_BAND_x, donde x es el número de la banda).

A_p : Es el factor aditivo de escalado específico por banda, obtenido del metadato (REFLECTANCE_ADD_BAND_x, donde x es el número de la banda).

Q_{cal} : Es el producto estándar cuantificado y calibrado para valores de pixel (ND).

Como variables de entrada a los modelos se utilizaron las bandas 2, 3, 4, 5, 6 y 7 del sensor OLI.

2.4 Algoritmos de clasificación supervisada

Las áreas de estudio fueron clasificadas utilizando los siguientes algoritmos: Máxima Verosimilitud (MLC), Support Vector Machine (SVM) y Random Forest (RF). Quedando conformadas nueve escenas clasificadas, (tres grupos de coberturas por tres algoritmos de clasificación). El primer algoritmo fue ejecutado utilizando el software ENVI 4.6., para SVM y RF se utilizaron rutinas compiladas en la caja de herramientas de Orfeo, presente en el software QGIS 2.14.

En relación a los parámetros, RF requiere de la definición del número de árboles de clasificación deseados (k) y el número de variables de predicción (m) a usar en cada nodo para hacer crecer el árbol. Estos valores fueron establecidos de forma empírica en 100 y 2 respectivamente. En el caso de SVM, el kernel utilizado fue del tipo Radial Basis Function. El par óptimo de los parámetros asociados: penalidad C y γ , se determinó mediante pruebas de validación cruzada.

2.5 Validación

Con el fin de determinar la precisión de los mapas resultantes se procedió con el cómputo de las matrices de confusión, utilizando como datos de validación la información derivada del relevamiento de los 70 lotes agrícolas de la campaña 2015 - 2016. Se calcularon los estadísticos Precisión Global (PG) y estadístico Kappa (κ) [11]. El coeficiente kappa (κ) computa el acuerdo entre la imagen clasificada y la verdad de terreno, debido únicamente a la exactitud de la clasificación, suprimiendo el acuerdo que cabría

esperar simplemente por azar. En [12] se propuso una escala para la interpretación del valor de dicho coeficiente donde valores menores a 0,40 se clasifican como pobres, 0,40-0,55 suficientes, 0,55-0,70 buenos, 0,70-0,85 muy buenos y mayores a 0,85 excelentes.

3 Resultados y discusión

En la Tabla 1 se exponen los valores del índice Kappa y las precisión global (en porcentaje), obtenidos al realizar la validación de los mapas. Adicionalmente los modelos de clasificación que hacían uso del grupo de entrenamiento G2, fueron evaluados considerando únicamente tipo de especie cultivada (modelos “2^a”).

Tabla 1. Estadísticos de validación para las clasificaciones con los distintos algoritmos y agrupaciones de clases de cobertura.

Grupo	Algoritmo	Modelo	Índice Kappa (κ)	Precisión Global (%)
G1	MLC	MLC_1	0,94	94,87
	SVM	SVM_1	0.81	82.82
	RF	RF_1	0.74	76.78
G2	MLC	MLC_2	0,99	99,21
	SVM	SVM_2	0.92	92.88
	RF	RF_2	0.71	74.93
	MLC	MLC_2 ^a	0.99	99.93
	SVM	SVM_2 ^a	1	100.00
	RF	RF_2 ^a	0.83	91.68
G3	MLC	MLC_3	0,92	96,29
	SVM	SVM_3	0.92	95.85
	RF	RF_3	0.88	94.37

Al analizar los estadísticos que validan las clasificaciones sobresale en primer lugar el excelente desempeño del algoritmo MLC en el mapeo de cubiertas agrícolas, lo cual se concuerda con los resultados obtenidos en anteriores estudios sobre la misma área geográfica [13, 14]. En este tipo de estudios, y en particular en la región bajo análisis, las amplias dimensiones de los lotes agrícolas facilita la obtención de abundantes muestras, representativas de cada clase, necesarias para un buen entrenamiento del algoritmo. Recientemente en otras regiones del globo, [15] utilizaron imágenes Landsat 8 para mapear las macro clases bosque, cultivos, suelo desnudo entre otras, y alcanzaron una precisión global del 90% al usar el algoritmo MLC. En [16], para el sur de Texas (USA), se utilizaron imágenes del satélite SPOT 5 en el mapeo de cultivos de verano

(maíz, algodón, sorgo, caña de azúcar y no cultivo), obteniendo con el algoritmo MLC una precisión global del 91%, concluyendo que es posible clasificar imágenes correspondientes a grandes áreas.

En el presente trabajo para el modelo que combina datos de referencia en el agrupamiento G3 y el algoritmo MLC las precisiones alcanzan el 99 por ciento con un índice κ de 0.99. Este excelente comportamiento puede tener su fundamento en el trabajo de campo previo, en el cual se relevaron lotes de maíz y soja en sus diversos estadios de crecimiento. Esto permitió la construcción de clases puras, para los conjuntos de entrenamiento y validación. Sobre los errores en los mapas temáticos [17] expresan que pueden no deberse a la fiabilidad del clasificador, sino ser el resultado de una mala localización de algunos elementos de los datos de prueba o de una identificación defectuosa de los atributos muestreados. Trabajando también en la pampa argentina y con cultivos similares, [18] concluyeron que la capacidad de discriminar entre las clases de coberturas de cultivos agrícolas está estrechamente relacionada con sus patrones fenológicos. Por ende clases de cultivos que posean similares patrones son a menudo mal clasificados. Este hecho fue identificado ya en los primeros estudios de mapeo realizados en la pampa Argentina, [19] encontró dificultades en la separación de las clases maíz y sorgo, especies pertenecientes a la misma familia botánica y de estructura morfológica similar. Estas afirmaciones corroboran la necesidad de un riguroso trabajo de campo y posteriormente a gabinete. No obstante pueden siempre existir errores en la discriminación debidos a la variabilidad causada por la heterogeneidad del suelo dentro de un lote, situaciones de estrés, climáticas, plagas o malezas, lo cual debe ser tenido en cuenta al momento de diseño del muestreo. En el área de estudio, para las clases Sorgo y Maní, no fue posible armar grupos por sus respectivos estadios fenológicos, debido al escaso número de lotes, que abarcan solo el 5% del área sembrada [20].

Cuando se analizan las clasificaciones realizadas empleando el algoritmo RF y las clases que representan los diferentes estadios de los cultivos “soja” y “maíz”, se aprecia un aumento en las precisiones logradas si son incluidas las clases sorgo y maní al modelo (RF_1). Estos resultados sugieren que la introducción de múltiples clases al problema de clasificación, permite al algoritmo la construcción de vías de clasificación más certeras. Sin embargo en este último modelo se observan fracciones de lotes de la clase soja tardía, mal clasificados como maní. Esta confusión podría deberse a la similitud en su reflectancia espectral [21]. En el modelo RF_3 el cual realiza una clasificación solo por las clases soja y maíz, se aprecian los mejores índices de validación para este algoritmo. En este modelo los datos de entrenamiento-validación para las especies soja y maíz en sus diversos estadios de crecimiento fueron unidos en las clases soja y maíz respectivamente, por lo cual el número de píxeles de entrenamiento por clase aumento de forma considerable. En [22] se infiere que cuanto más se incrementan los datos de entrenamiento, mayor es la precisión del clasificador, de forma paralela hay menos necesidad de ajustar los parámetros del algoritmo RF. Por otro lado [23] señalan que muchos métodos de discriminación incluidos los derivados del machine learning, a menudo son más precisos y eficientes cuando trabajan solamente con dos clases. Por otro lado cuando los resultados de la clasificación por el modelo RF_2, son unidos por su respectiva especie (es decir “soja” y “maíz”) y se recalculan los estadísticos de validación (RF_2^a), se aprecia que estos son menores a los alcanzados por el modelo RF_3. Este comportamiento puede relacionarse a lo encontrado por [24] quienes trabajando con datos multitemporales de Landsat 5 en la provincia de Granada, España, observaron

que el error respecto al modelo calibrado, superaba el 5 %, con disminuciones de mayores al 70% del conjunto de datos de entrenamiento. Este comportamiento evidencia que RF es sensible a la reducción del conjunto de entrenamiento.

Por último, al examinar los índices de precisión de los mapas generados con el algoritmo SVM, sobresale en primera instancia la capacidad del algoritmo para mantener buenos índices respecto a RF, con los diversos grupos de entrenamiento-validación. Estos convenientes valores pueden relacionarse a la capacidad del algoritmo para detectar y superar las bandas ruidosas, sumado al gran impacto de la optimización de los parámetros C y γ en la precisión de la clasificación, la cual permite sortear el sobreentrenamiento y una pobre generalización [25]. Sin embargo a diferencia de lo ocurrido en RF y de manera semejante a la prueba con el algoritmo MLC, la precisión del mapa resultante también se incrementa cuando en el modelo se analizan únicamente las clases correspondientes a las especies “soja” y “maíz” en sus diferentes estadios fenológicos (SVM_2). Cuando el objetivo es generar mapas de superficies destinadas a los cultivos de “soja” y “maíz”, sin diferenciar por momento de siembra, los resultados manifiestan que se obtienen mejores productos si al modelo de clasificación son ingresadas las clases que representan los diversos estadios fenológicos de los cultivos, y luego se combinan los píxeles resultantes de cada especie (SVM_2^a). En la mayoría de las aplicaciones, el éxito del aprendizaje automático está fuertemente afectado por la calidad de los datos (información redundante, ruidosa o poco confiable). En este sentido haber establecido las clases de acuerdo a los respectivos estadios fenológicos pudo eliminar los valores atípicos, y de esta forma evitar el sobreentrenamiento, permitiendo alcanzar un excelente desempeño en la fase de prueba [26].

En la Figura 2 se exponen recortes de los mapas resultantes para los modelos MLC 1, 2, y 3 respectivamente, con la correspondiente área en color real.

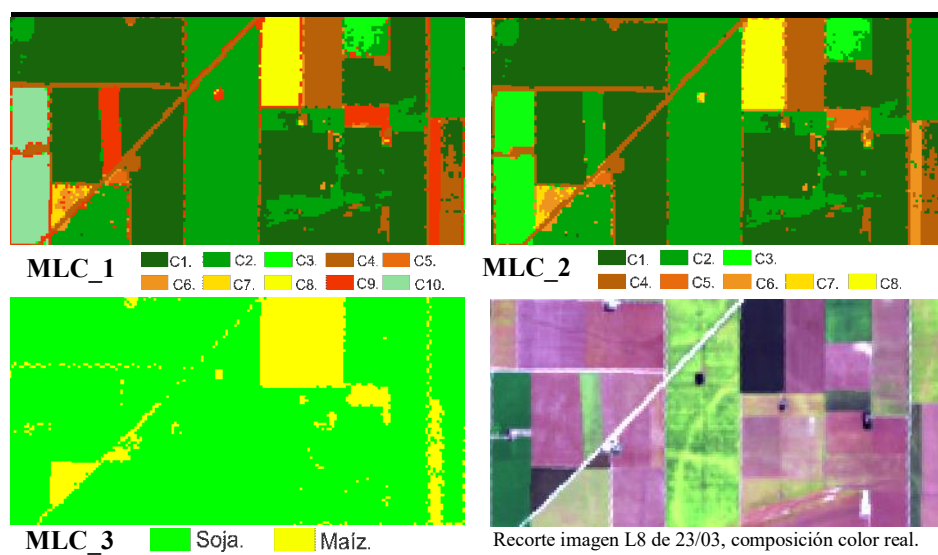


Figura 2. Recorte de imágenes clasificadas con el algoritmo MLC, correspondientes a los grupos de entrenamiento G1, G2, y G3, y recorte de la misma área en composición color real.

Respecto a las deficiencias de cada enfoque de clasificación se puede mencionar, que debido a la característica paramétrica del clasificador de máxima verosimilitud, las altas precisiones alcanzadas en este estudio dependieron considerablemente del tamaño y confiabilidad del conjunto de entrenamiento. En la práctica, se requiere un tamaño mucho más grande que el mínimo ideal para representar completamente la distribución en un espacio multidimensional. La escasez de datos de verdad de terreno para el entrenamiento es el factor más limitante en el mapeo de los cultivos usando el enfoque MLC [2]. Respecto a los algoritmos SVM y RF se subraya la necesidad de ajustar por medio de una búsqueda empírica, los mejores valores de los parámetros para el dominio donde están siendo utilizados. La bibliografía en el tema brinda aproximaciones, sin embargo no se sabe de antemano cuales son los preferibles para un problema dado [27]. En este sentido SVM es más complejo que RF [28]. Por último en lo que respecta a la demanda computacional se destaca los requerimientos por parte de SVM, en lo referente a la optimización de los valores de penalidad C y del parámetro γ . Este costo aumenta considerablemente con el número de clases y tamaño del conjunto de entrenamiento. Respecto a los tiempos de entrenamiento, en [29] cuando se duplicaron el tamaño del conjunto de entrenamiento a ingresar al modelo SVM, el tiempo de entrenamiento aumento en una proporción mayor.

Por último es necesario recalcar que la fortaleza y potencialidad de los métodos aquí presentados se sustentan en características inherentes al problema en estudio. Se menciona, la dimensión de los establecimientos en el área analizada (media=100ha, permitiendo eliminar los píxeles fronterizos), el reducido número de especies cultivadas, y el contraste de las principales familias botánicas. También es significativo la disponibilidad de datos espectrales clave, convenientes resoluciones espaciales y temporales. En este sentido es imprescindible disponer de imágenes libres de nubes cuando la diferenciación espectral entre los tipos de cobertura se maximiza. Estas consideraciones posibilitan la elaboración de mapas precisos y exactos, con antelación a la finalización de las campañas, lo cual es beneficioso para la planificación, estimación y generación de presupuestos.

4 Conclusiones

Los resultados que se obtuvieron en el presente estudio permiten afirmar que las imágenes satelitales, posibilitan el mapeo de la diversidad de coberturas presentes en el área central de la provincia de Córdoba, con anterioridad a la finalización de sus ciclos.

Las mejores precisiones se alcanzaron al utilizar grupos de entrenamiento y validación cuyas clases se limitaban a un estadio específico del ciclo de cada cultivo. La metodología usada, que incluyó un riguroso trabajo de campo, en lo que respecta a la recolección de muestras representativas de los tipos de cobertura de interés sobre el área de estudio, y posteriormente a gabinete, la determinación de clases de entrenamiento-validación puras fue base para que los algoritmos de clasificación lograron discriminar, incluso, clases de cobertura con similares patrones espectrales.

Finalmente, por la simplicidad y sencillez de su ejecución, junto a su bajo costo computacional se concluye como más conveniente el uso del algoritmo de Máxima

Verosimilitud en el mapeo de cultivos agrícolas. Este algoritmo alcanzó excelentes resultados utilizando una única imagen de fin de temporada, y un conjunto de datos de referencia a campo que englobaba, de forma precisa, la heterogeneidad de las coberturas presentes en el área.

Agradecimientos:

Los autores expresan su gratitud a la Secretaría de Ciencia y Técnica – Universidad Nacional de Córdoba - Argentina (Secyt–UNC) por el apoyo financiero que permitió la realización de este trabajo.

Referencias

1. Matton, N., Canto, G.S., Waldner, F., Valero, S., Morin, D., Inglada, J., Arias, M., Bontemps, S., Koetz, B., Defourny, P.: An automated method for annual cropland mapping along the season for various globally-distributed agrosystems using high spatial and temporal resolution time series. *Remote Sens.* 7, 13208–13232 (2015).
2. Zhong, L.: Efficient crop type mapping based on remote sensing in the Central Valley, California, (2012).
3. Naylor, R.: Expanding the boundaries of agricultural development. *Food Secur.* 3, 233–251 (2011).
4. Vapnik, V.N.: *Statistics for Engineering and Information Science*. Springer, New York (2000).
5. Breiman, L.E.O.: Random Forests. *Mach. Learn.* 45, 5–32 (2001).
6. Daza, C.G.: *Zonas Agroecológicas Homogéneas Córdoba*. Instituto Nacional de Tecnología Agropecuaria (2009).
7. Cabrera, A.L.: *Regiones fitogeográficas argentinas*, (1976).
8. Gorgas, J.A., Tassile, J.L., Jarsún, B., Zamora, E., Bosnero, H., Lovera, E., Ravelo, A.: Recursos naturales de la provincia de Córdoba, los suelos: nivel de reconocimiento escala 1: 500.000. Agencia Córdoba DAC y TSEM Dirección Ambiente y Instituto Nacional de Tecnología Agropecuaria INTA-Manfredi (2003).
9. Ministerio de Agroindustria de la Nación: Datos abiertos de agroindustria, <https://datos.magyp.gob.ar/reportes.php?reporte=Estimaciones>.
10. Ariza, A.: Descripción y Corrección de Productos Landsat 8 LDCM (Landsat Data Continuity Mission), (2013).
11. Foody, G.M.: Status of land cover classification accuracy assessment. *Remote Sens. Environ.* 80, 185–201 (2002).
12. Monserud, R.A., Leemans, R.: Comparing global vegetation maps with the Kappa statistic. *Ecol. Modell.* 62, 275–293 (1992).
13. Willington, E., Nolasco, M.: Clasificación supervisada de suelos de uso agrícola en la zona central de Córdoba (Argentina): comparación de distintos algoritmos sobre imágenes Landsat. Presented at the (2013).
14. Nolasco, M., Willington, E., Bocco, M.: Uso del suelo agrícola: comparación entre series temporales e imágenes satelitales individuales para su clasificación. *Rev. Investig. la Fac. Ciencias Agrar.* 26, 017–021 (2016).
15. Jia, K., Wei, X., Gu, X., Yao, Y., Xie, X., Li, B.: Land cover classification using Landsat 8 Operational Land Imager data in Beijing, China. *Geocarto Int.* 29, 941–951 (2014).
16. Yang, C., Everitt, J.H., Murden, D.: Evaluating high resolution SPOT 5 satellite imagery for crop identification. *Comput. Electron. Agric.* 75, 347–354 (2011).
17. Tso, B., Mather, P.M.: *Classification methods for remotely sensed data*. CRC press, Boca Raton (2009).
18. Guerschman, J.P., Paruelo, J.M., Bella, C. Di, Giallorenzi, M.C., Pacin, F.: Land cover classification in the Argentine Pampas using multi-temporal Landsat TM data. *Int. J. Remote Sens.* 24, 3381–3402 (2003).
19. Badhwar, G.D.: Landsat Classification of Argentina Summer Crops. *Remote Sens. Environ.* 21, 111–117 (1987).
20. Bolsa de cereales de Córdoba: Datos finales de producción, <http://www.bccba.com.ar/>.
21. Schultz, B., Immitzer, M., Formaggio, A.R., Sanches, I.D.A., Luiz, A.J.B., Atzberger, C.: Self-guided segmentation and classification of multi-temporal Landsat 8 images for crop type mapping in Southeastern Brazil. *Remote Sens.* 7, 14482–14508 (2015).
22. Tatsumi, K., Yamashiki, Y., Angel, M., Torres, C., Leonidas, C., Taípe, R.: Crop classification of upland fields using Random forest of time-series Landsat 7 ETM + data. *Comput. Electron. Agric.* 115, 171–179 (2015).

23. Camps-Valls, G., Gómez-Chova, L., Calpe-Maravilla, J., Soria-Olivas, E., Martín-Guerrero, J.D., Moreno, J.: Support vector machines for crop classification using hyperspectral data. In: Goos, G., Hartmanis, J., and van Leeuwen, J. (eds.) *Iberian Conference on Pattern Recognition and Image Analysis*. pp. 134–141. Springer (2003).
24. Rodríguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M., Rigol-Sanchez, J.P.: An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS J. Photogramm. Remote Sens.* 67, 93–104 (2012).
25. Foody, G.M., Mathur, J.: A relative evaluation of multiclass image classification by support vector machines. *IEEE Trans. Geosci. Rem. Sens.* 42, 1–9 (2004).
26. Camps-Valls, G., Gómez-Chova, L., Calpe-Maravilla, J., Martín-Guerrero, J.D., Soria-Olivas, E., Alonso-Chordá, L., Moreno, J.: Robust support vector method for hyperspectral data classification and knowledge discovery. *IEEE Trans. Geosci. Remote Sens.* 42, 1530–1542 (2004).
27. Hsu, C.W., Chang, C.C., Lin, C.J.: *A practical guide to support vector classification*, (2003).
28. Rodríguez-Galiano, V.F., Chica-Rivas, M.: Evaluation of different machine learning methods for land cover mapping of a Mediterranean area using multi-seasonal Landsat images and Digital Terrain Models. *Int. J. Digit. Earth.* 1–18 (2012).
29. Huang, C., Davis, L., Townshend, J.: An assessment of support vector machines for land cover classification. *Int. J. Remote Sensing.* 23, 725–749 (2002).